Chapter 3

Hybrid Deep Learning Segmentation Models for Atherosclerotic Plaque in Internal Carotid Artery B-mode Ultrasound

Summary

The automated and accurate carotid plaque segmentation in B-mode ultrasound (US) is an essential part of stroke risk stratification. Previous segmented methods used AtheroEdge™ 2.0 (AtheroPoint™, Roseville, CA) for the common carotid artery (CCA). This study focuses on automated plaque segmentation in the internal carotid artery (ICA) using solo deep learning (SDL) and hybrid deep learning (HDL) models.

The methodology consists of a novel design of 10 types of SDL/HDL models (AtheroEdge™ 3.0 systems (AtheroPoint™, Roseville, CA) with a depth of four layers each. Five of the models use cross-entropy (CE)-loss, and the other five models use Dice similarity coefficient (DSC)-loss functions derived from UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+. The K10 protocol (Train:Test::90%:10%) was applied for all 10 models for training and predicting (segmenting) the plaque region, which was then quantified to compute the plaque area in mm^2. Further, the data augmentation effect was analysed.

The database consisted of **970** ICA B-mode US scans taken from **99** moderate to high-risk patients. Using the difference area threshold of 10 mm^2 between ground truth (GT) and artificial intelligence (AI), the area under the curve (AUC) values were **0.91**, **0.911**, **0.908**, **0.905**, and **0.898**, all with a *p*-value of <0.001 (for CE-loss models) and **0.883**, **0.889**, **0.905**, **0.889**, and **0.907**, all with a *p*-value of <0.001 (for DSC-loss models). The correlations between the AI-based plaque area and GT plaque area were **0.98**, **0.96**, **0.97**, **0.98**, and **0.97**, all with a p-value of <0.001 (for CE-loss models) and **0.98**, **0.98**, **0.97**, **0.98**, and **0.98** (for DSC-loss models). Overall, the online system performs plaque segmentation in less than one second. We validate our hypothesis that HDL and SDL models demonstrate comparable performance. SegNet-UNet was the best-performing hybrid architecture.

3.1 Introduction

Atherosclerotic plaques in the internal carotid artery (ICA) may rupture, thus embolizing the brain blood vessels and leading to a stroke [122, 123]. However, only a minority of plaques are unstable and rupture, resulting in an annual stroke rate of 1-2% in asymptomatic patients with greater than 80% stenosis. Thus, operating on all patients with >80% stenosis will result in many preventable and needless surgeries. In addition, such surgeries are associated with a 3% preoperative stroke rate. These plaques can be characterized for stroke risk and, thus, require automated plaque segmentation, quantification, and monitoring over time [124].

Plaque imaging uses several medical imaging modalities such as magnetic resonance imaging (MRI) [24], computed tomography (CT) [125, 126], X-ray [127], and ultrasound (US) [8, 128]. US offers several benefits such as non-invasive, ergonomic, economic, radiation-free, portable, and real-time evaluation. It is, therefore, preferred over MRI and CT. Further, US offers features like harmonic and compound imaging that provides noise cancellation during beam formations [59, 129], thus offering high-resolution imaging with 0.0625 mm/pixel. Due to the above benefits, US has been adapted for the diagnosis of several diseases such as characterizing breast cancer lesions [130], liver cancer [131, 132], carotid artery disease [133, 134], thyroid cancer [135, 136], coronary artery disease [137], and cardiovascular risk [134]. One such application of US is plaque tissue characterization (PTC) in humans' internal carotid arteries (ICAs). PTC is necessary to understand plaque vulnerability or plaque symptoms. Researchers have conducted several studies to improve the general understanding of the *symptomatology of the plaque* in (i) carotid arteries using CT [48, 126, 138], (ii) coronaries using OCT [121], and in (iii) carotid arteries using B-mode US based on machine learning [36, 139-141] or deep learning (DL) [142, 143] paradigms. Even though the above studies demonstrated the ability to characterize/classify the plaque type, they used manual or semi-automated delineation methods for the *segmentation* of plaque before being used for PTC.

Several methods have been proposed for the segmentation of “plaque” or “plaque along with the walls” in the common carotid artery (CCA) in the last decade [144, 145]. These methods mainly adapted (a) geometric feature extraction, line fitting, and classification [146] or (b) scale-space paradigm [40, 147], mainly applied for US scans for low- to moderate plaque. These conventional methods are challenging since they process grayscale intensity in local or global frameworks [148].

Artificial intelligence (AI) has started to utilize DL in healthcare applications, particularly in radiology [93, 149]. Due to the superior ability of DL solutions—particularly automated feature extraction derived from the knowledge of cohort for segmentation or classification [150] we hypothesize that DL models are superior to conventional models [148, 151] for plaque segmentation, especially in ICA. Thus, the objective of this study is focused on an automated plaque segmentation in B-mode ICA US using either solo deep learning (SDL) models or hybrid deep learning (HDL) models that are characterized by the fusion of two DL models. Recently, a study investigated plaque segmentation and area measurements using DL in CCA [107, 133]. Even though the authors used DL for plaque wall segmentation, they did not attempt the same for ICA. The UNet++ ensemble model was attempted for CCA/ICA plaque segmentation but was not fully automated and did not use the power of hybrid paradigms [98].

This study focuses on completely automated plaque detection (or segmentation) and quantification in moderate- to high-risk patients using B-mode US ICA. Thus, the objective of this study is to evaluate the plaque burden in ICA rather than evaluating carotid intima-media thickness (cIMT).

This study is the first of its kind in that it segments the far wall plaque in ICA. There are inherent challenges when segmenting plaque in ICA, unlike in CCA. It includes (a) variation in the grayscale intensities such plaque echolucency, where hypo-intensity distribution is between two hyper-intensity plaques and vice-versa [27, 152]; (b) shadows in the far wall ICA plaque region due to the presence of calcium in the near-wall; (c) sharp changes in plaque burden due to inherent geometry of ICA and sinus bulb; and (d) a relatively larger plaque area due to the presence of plaque components such as calcium, lipids, fibrin, and fibrosis [153, 154] in high-risk patients (see Figure 3.1). Thus, there is a clinical need to automatically segment and quantify far-wall plaques in ICA.

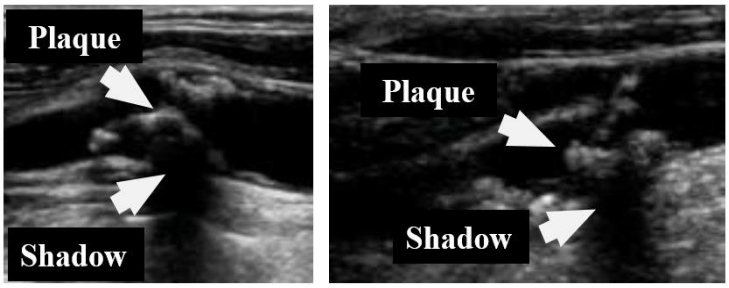


Figure 3.1 Example of B-mode ICA US scans showing large plaque with black shadow.

Recently published studies have confirmed that HDL is superior to SDL for segmentation applications [155, 156]. Our spirit of HDL stems from these studies. Thus, this study uses SDL and HDL models for plaque with wall segmentation in the B-mode US ICA scans and, further, automatically quantifies the plaque area (which is an indicator for stroke/CVD risk assessment). This study does not explicitly focus on the "near-wall region" analysis, and so this topic will not be discussed here. Figure 3.2 shows the global system for segmentation of plaque in the ICA using SDL and HDL models. The three significant subsystems shown in Figure 3.2 are (i) plaque and wall segmentation, (ii) quantification of the plaque segmentation region, and (iii) performance evaluation of the DL/HDL-based area given the ground truth (GT) area. The plaque wall segmentation consists of six DL models (3 CE-loss-based and 3 DSC-loss-based) and four HDL models (2 CE-loss-based and 2 DSC-loss-based). These 10 models were derived using three SDL models (UNet, UNet+, and SegNet) and two HDL models (SegNet-UNet, and SegNet-UNet+).

The second stage consists of the quantification subsystem of the plaque-wall area (in mm^2) given the binary ICA plaque segmented masked regions. The estimated AI-based area was then compared against the GT area of the plaque envelope as traced by a clinically (LS) trained scientist. Finally, the performance evaluation system evaluates the SDL/HDL system, which compares the AI-based area against the GT area to compute the receiving operating curve (ROC), coefficient of correlations (CC), Bland-Altman's (BA) plot, cumulative distribution plots (CDP) of area error, and mean absolute

difference between GT area and AI-based area. All 10 models were evaluated in sequence, and the mean performance was computed for all models.

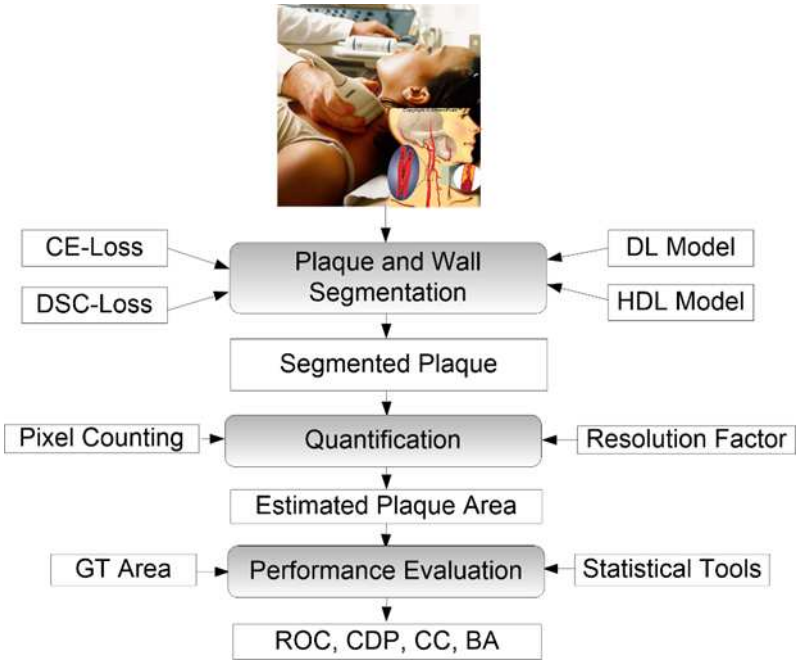


Figure 3.2 AtheroEdge™ 3.0 DL, HDL, ROC, CDP, CC, BA, CE-loss, and DSC-loss (see acronym table).

This study offers the following novel components: (i) design of HDL models for plaque segmentation and its quantification in ICA B-mode US scans; (ii) comparative approach of four HDL models against six SDL models for ICA plaque segmentation; (iii) analysis of 10 performance evaluation attributes to demonstrate the ability of CE-loss superior to DSC-loss functions in HDL/SDL framework; and (iv) justification of image frame selection in the ultrasound video by using the combinations of (a) the interval between selected frames during data preparation and (b) standardized five augmentation protocols.

3.2 Methodology: Data Preparation, AI Architectures, and Experimental Protocol

The main components of the methodology section consist of three main parts: Subsection 3.2.1 discusses patient selection, data collection, and data preparation. Subsection 3.2.2 explains the architectures for SDL and HDL models, along with their loss functions. Finally, subsection 3.2.3 presents the experimental protocol used for all 10 DL models.

3.2.1 Patient cohort, data collection, and data preparation

3.2.1.1 Patient Cohort

As part of the cross-section study, 99 B-mode US videos of left and right ICA were acquired from 99 patients at Imperial College London UK. These patients were either symptomatic or asymptomatic, having carotid stenosis > 50%. The baseline characteristics of patients included 47M/52F and a mean age of 75.04±9.96 years, ranging from 45 years to 96 years. No clinical information other than the ultrasound videos was available for this cohort.

3.2.1.2 Carotid Ultrasound Image Acquisition

Philips ultrasound system iU22 (Philips Ultrasound, Bothell, USA) equipped with a linear transducer (L9-3) of 10 MHz resonant frequency was used for carotid artery examination. An expert neurologist with 15 years of experience acquired 99 ICAB-mode US from both the left and right ICA, taken from 99 patients. The carotid arteries of study participants were examined in the supine position with the head tilted backward. Image acquisition follows a two-step protocol: (i) identifying of carotid arteries using transverse scanning and (ii) capturing the images of the walls of carotid arteries by an orthogonal rotation (i.e., 90 degrees) of the ultrasound probe and moving the probe to ICA position. A six-second video was collected for the ICA segment [157]. The average image resolution factor was 0.0625 mm/pixel. The image acquisition protocol followed the recommendations provided by the American Society of Echocardiography Task Force (ASE) [158] and was discussed in detail in our previous studies [40, 147, 159, 160]. In their recent clinical paper, Johri *et al.* [160] proposed recommendations for computing PA or TPA measurements in US scans.

3.2.1.3 Data Preparation

The US videos were divided into low-plaque and high-plaque categories based on a vascular surgeon's input, which was based on the visual judgment and experience of 15 years. Because two US video scans were of poor image quality, as recommended by the vascular surgeon (AAG), we decided to exclude those from experiments, reducing our dataset to 97 US videos. Further, each US video was transformed into still image frames for further processing. A selection of 10 US frames (at the interval of 10) was made from each US video; thus, a pool of 970 image frames was prepared as that final dataset used for experiments. The justification of the interval of 10 is presented in the discussion section, where we relax this assumption by choosing no interval and, instead, using the augmentation of the data sets. Figure 3.3 shows two high-plaque and two low-plaque grayscale images (top-row) and their corresponding GT binary mask images (bottom row).

3.2.1.4 Ground Truth Data Preparation

A scientist trained by 20 years' experience radiologist prepared the GT binary mask by tracing the envelope around the plaque and along the lumen intima (LI) and media adventitia (MA) borders of the ICA far wall [23]. The tracings of the envelope explicitly used ImgTracer™ (AtheroPoint™, Roseville, CA, USA) [161, 162]. The same binary masks were used as GT for all SDL/HDL systems in this study. Due to the cost and tedious nature of the manual tracings, only one tracer was used for the current study.

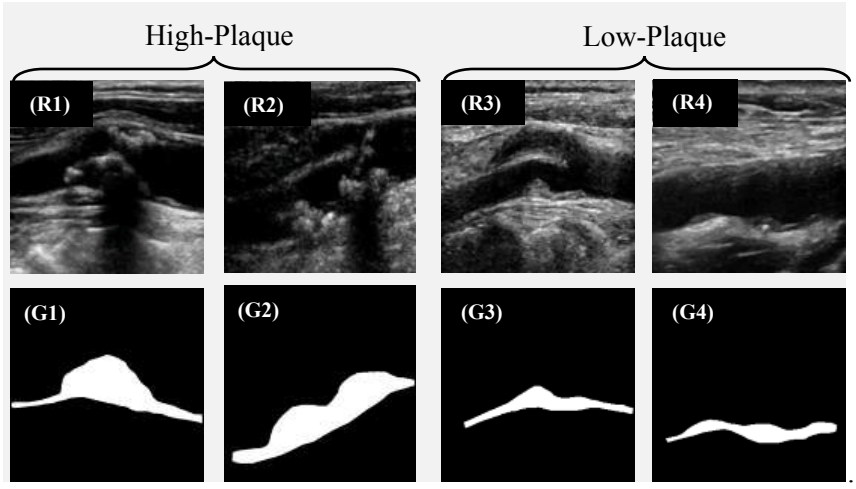


Figure 3.3 (a) Raw and binary GT masks. R1 and R2 (high-plaque) and R3 and R4 (low-plaque) represent the raw ICA B-mode scans. Images marked G1, G2, G3, and G4 are the corresponding GT binary masks.

3.2.2 Architecture of SDL and HDL models

Five segmentation models, namely UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+ under CE-loss and DSC-loss conditions, were adapted for our experimental protocol. Thus, we first briefly study the SDL models, followed by HDL models.

3.2.2.1 Architecture of SDL Models: UNet, UNet+, SegNet

UNet Architecture

This is one of the most popular image segmentation methods and was initially proposed by Ronneberger *et al.* [163]. This architecture used the concept of deconvolution, which was introduced by Zeiler *et al.* [164]. Figure 3.4 (a) below represents the UNet architecture consisting of the encoder-decoder combination having a depth of four. The 'U' shaped model-is connected by a bridge network (in the bottom of the U shape consisting of 3x3x124 filters, red in colour, and ReLU, green in colour). The encoder comprises a combination of convolution layers followed by ReLU and MaxPooling layers, whereas the decoder consists of up-convolution, followed by convolution, ReLU, and MaxPooling layers. Each encoder stage uses a convolutional filter with a size of 3x3 and MaxPooling, which down-samples the image for the next stage. After the bridge network, the decoder network up-samples the image with 2x2 up-convolutional layers. Each encoder stage supplies the spatial features to the corresponding decoder stage via a skip connection [165]. Thus, the decoder's output is updated with spatial information from the corresponding encoder stage. Finally, after the fourth decoder stage, an efficient ADAM optimizer reduces the loss function and a SoftMax classifier that classifies the up-sampled data into two classes: ICA plaque and background.

UNet+ Architecture

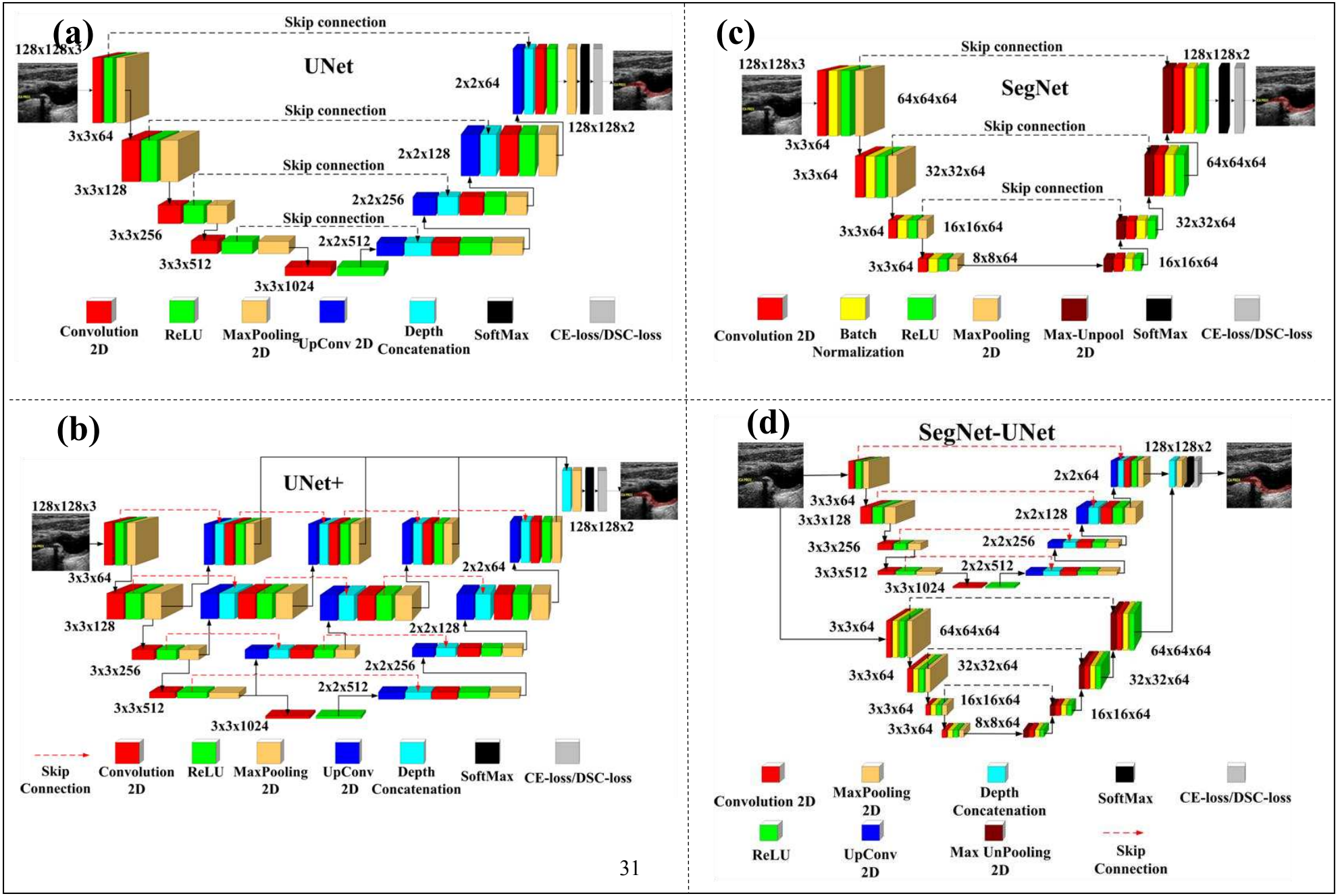
Another variant of UNet architecture used for ICA plaque segmentation is UNet+ (shown in Figure 3.4 (b)). The architecture has some intermediate encoder stages between compression and expansion stages. For depth, d=4 (encoder and decoder) stages of UNet+, the first intermediate stage has d-1 (4-1=3) first intermediate encoder stages, d-2 (4-2=2) second intermediate stages, and d-3 (4-3=1) one last intermediate stage. Thus, the given UNet+ architecture has four encoder stages: three first encoder stages, two second intermediate stages, one last third intermediate stage, and four decoder stages. The solo UNet model has a limitation of the optimal depth of the encoder-decoder network. The depth of the network depends on the types of applications and the amount of training data available. These limitations are overcome by applying varying depths of intermediate up-sampling units inside the architecture. The previous intermediate up-sampling stage connects each decoder stage to the same resolution through redesigned skip connection. Therefore, UNet+ can be regarded as the amalgamation of UNet's different depths in its architecture. UNet+ avoids the choice of network depths. UNet+ after the fourth decoder stage uses an efficient ADAM optimizer that reduces the loss function and a SoftMax classifier that classifies the up-sampled data into two classes: ICA plaque and background.

SegNet Architecture

This is a semantic segmentation network, unlike UNet. However, it uses approximately 10 times fewer learning parameters due to its architecture. Our SegNet architecture comprises four depths of encoders and decoder blocks and has no bridge network to join them as shown in Figure 3.4 (c). Each encoder block consists of a convolutional layer, batch normalization, ReLU, and MaxPooling layers. The convolutional filter size is 3x3 throughout the system in all networks. However, the depth or number of filters is changed in each stage to extract more features maps than the previous layer.

3.2.2.2 Architecture of Hybrid DL Models: SegNet-UNet and SegNet-UNet+

The concept of HDL was introduced when Szegedy *et al.* [166] first fusion design by fusing Inception with ResNet. In our study, the results of UNet, UNet+, and SegNet architectures were encouraging, which led us to establish two hybrid architectures using the combinations of SDL models. Figures 3.4 (d) and 3.4 (e) show the architectures of two hybrid models. Both hybrid models are parallel addition of SegNet with UNet and UNet+. We fused the SegNet with UNet, and developed SegNet-UNet. Similarly, SegNet fused with UNet+ yielded SegNet-UNet+ hybrid models. These hybrid models share the input layer, SoftMax layer, and loss function layer. Lastly, we tested the two HDL models with the same hyper-parameter conditions and two-loss functions (CE-loss and DSC-loss), discussed as follows.


(a)
UNet
Skip connection
128x128x3
3x3x64
3x3x128
3x3x256
3x3x512
3x3x1024
2x2x64
2x2x128
2x2x256
2x2x512
128x128x2
Convolution 2D
ReLU
MaxPooling 2D
UpConv 2D
Depth Concatenation
SoftMax
CE-loss/DSC-loss
(c)
SegNet
64x64x64
32x32x64
16x16x64
8x8x64
Batch Normalization
Max-Unpool 2D
(b)
UNet+
Skip Connection
(d)
SegNet-UNet
Max UnPooling 2D

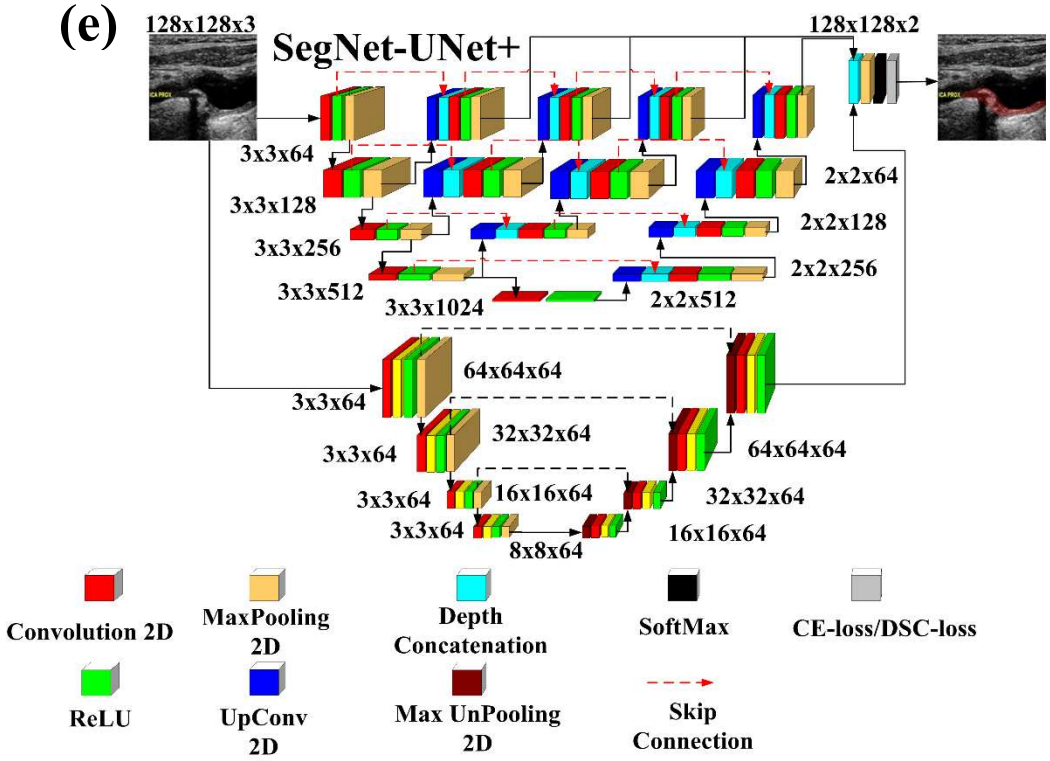


Figure 3.4 Three SDL and two basic HDL segmentation models.

3.2.3 Loss Functions for SDL and HDL models

Our system used loss function, namely CE-loss and DSC-loss, in SDL/HDL models, which correspond to two different types of models in this paper. The CE-loss function, labeled as L_{CE} for all the models, can be mathematically given as Eq. 3.1:

$$L_{CE} = -[(y_i \times \log a_i) + (1-y_i) \times \log(1-a_i)] \quad (3.1)$$

Where y_i is the input GT label 1, $(1-y_i)$ is GT label 0, and a_i represents the SoftMax classifier probability. The DSC-loss function is based on the Sørensen-Dice similarity coefficient for measuring the amount of overlap between two segmented images. The generalized Dice loss function L_{DSC} between the ground truth image Y and the corresponding AI-estimated image $\hat{Y}$ is mathematically represented by Eq. 3.2 [167]:

$$L_{DSC} = 1 - \left[\left(\frac{2\times\sum_{p=1}^{P} Y_p \times \hat{Y}_p}{\sum_{p=1}^{P} Y_p^2 \times \hat{Y}_p^2}\right) + \left(\frac{2\times\sum_{p=1}^{P}(1-Y_p)\times(1-\hat{Y}_p)}{\sum_{p=1}^{P}(2-Y_p^2 \times \hat{Y}_p^2)}\right)\right] \quad (3.2)$$

Where, P is the total number of elements in the x, y-direction of the image. The quantity inside a square bracket represents the Dice similarity coefficient for binary segmentation. Note that for any input model (SDL or HDL), the number of layers, their organization, the hyper-parameters—such as learning rate,

number of epochs, batch size, and optimization—are kept constant for both loss functions. The tabular representation of these hyper-parameters will be discussed in the next section.

3.3 Experimental Protocol

3.3.1 Cross-Validation

A total of 970 images are used in our system. A well-known data partitioning method (K10) was used in our system. Of these images, 90% (873 images) are used for training and 10% (97 images) are used for testing. An internal validation system is also used in our system with training data. Training data is further divided into 90% training (786 images) and 10% internal validation (87 images) data. However, internal validation can be biased because the images are known to the system in each epoch. Therefore, we focused only on test images that are entirely unknown to the system. Also, we expect the system to be unbiased and optimized for the test data.

Further, to make our system more reliable, we added one more layer of accuracy. We returned all 10% (97 images) of the test data back into the main pool, fetched a new and unique 10% portion of the data, and repeated the experiment. The exact process was repeated for 10 unique partitions to train and test for the cohort. Note that, in each combination for K10, the test images were unknown to the system. In the results section, we have presented and discussed the mean of all 10 partitions. For each test patient, the binary mask was generated, and the plaque area was quantified. Figure 3.5 shows example images of GT binary mask in the first row and corresponding estimated binary mask from UNet model with CE-loss.

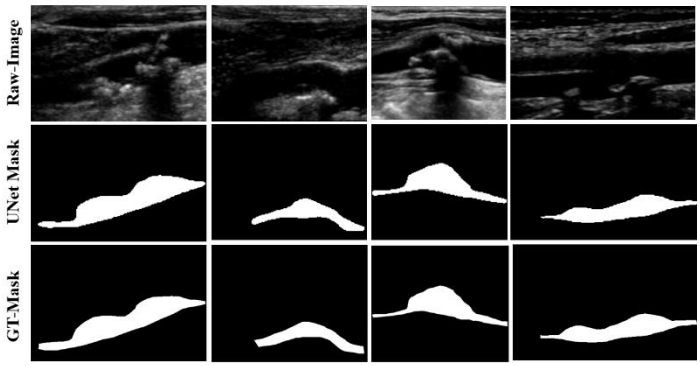


Figure 3.5 Raw grayscale ICA image (top), AI-estimated binary mask using UNet with CE-loss function (middle), and GT binary mask (bottom).

3.3.2 Plaque Quantification

There are two-fold advantages of computing the ICA carotid plaque area. First, the automated carotid plaque area can be used to study the relationship with CVD biomarkers [168], and (ii) the carotid

plaque area can be used to evaluate the performance of the segmentation model by comparing AI-based area against GT-area [43]. Thus, our main objective is to evaluate the AI-area against the GT-area. The training models generate an AI-based binary mask over the far wall plaque area corresponding to the test images. The plaque area is measured by counting the number of white pixels in the far wall plaque region and converting this value into mm dimensions using a resolution factor of 0.0625 mm to a pixel. If N represents the total number of images in the database, $\bar{\mathcal{A}}_{ai}(m)$ represents the mean plaque area for "ai" model "m", $\bar{\mathcal{A}}_{gt}$ represents the corresponding mean area of the GT binary mask, and the mean areas can be mathematically calculated using Eq. 3.3:

$$\bar{\mathcal{A}}_{ai}(m) = \frac{\sum_{n=1}^{N} \mathcal{A}_{ai}(m,n)}{N} \;\&\; \bar{\mathcal{A}}_{gt} = \frac{\sum_{n=1}^{N} \mathcal{A}_{gt}(n)}{N} \qquad (3.3)$$

3.4 Experimental Results

3.4.1 Hyper-parameters optimization and performance

The hyper-parameter conditions for all five DL models are shown in Table 3.1. Training is performed for 90% data for 100 epochs, and the training weights (models) are saved for the test data evaluation. The experiment with one DL model is performed for 10 unique partitions of data. In each partition, a new set of test data is presented to the DL system, and the mean of classification parameters for 10 partitions are shown in Tables 3.2 and 3.3. The same experiment is repeated for CE-loss and DSC-loss, with the results presented in Tables 3.2 and 3.3, respectively.

3.4.2 Plaque Segmentation and Visual Representations

Figure 3.6 shows a zoomed-in version of the AI-estimated area in red and the difference between GT and the AI-estimated area in green. The two images represent moderate plaque and high plaque images from the database. Figures 3.7 (a) and (b) represent the AI-estimated plaque area and the difference plaque area in high-plaque conditions, while Figures 3.7 (c) and (d) represent the same for low-plaque images, respectively. The first row represents the raw image and green overlay of the binary mask over the raw image. The second row represents the AI-estimated plaque area (PA) and absolute difference plaque area (ADPA) with the UNet model. The two columns on the left represent PA and ADPA for DL with CE-loss; the two columns on the right represent the same for DL with DSC-loss. In the same figure, rows 2, 3, and 4 represent AI-estimated plaque area and difference PA by SDL (i.e., UNet, UNet+, SegNet). Meanwhile, rows 5 and 6 illustrate the same for SegNet-UNet and SegNet-UNet+, respectively.

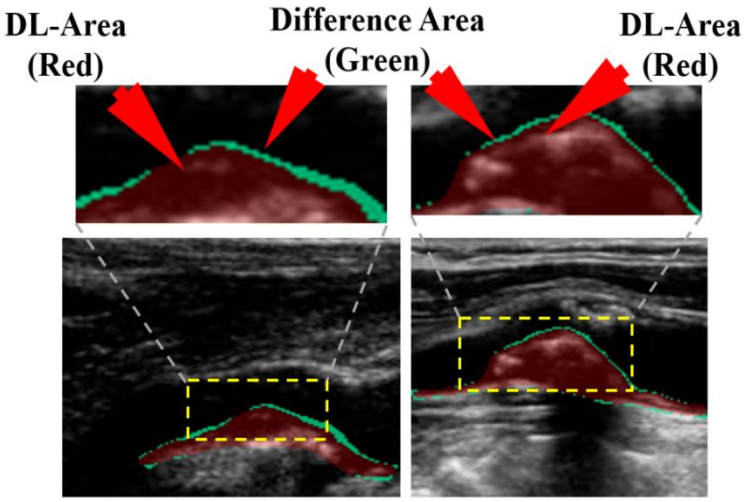


Figure 3.6 Examples showing plaque segmentation in B-mode ICA and B-mode US images. Left: moderate-plaque. Right: high-plaque. The AI-based plaque region is shown in red, and the difference between AI and GT is shown in green colour. The yellow dotted regions show high curvature zones. Left image: A_{gt}=48.79 mm^2 A_{ai}=43.79 mm^2; (A_{gt}-A_{ai})=5.22 mm^2; Right image: A_{gt}=99.22 mm^2 A_{ai}=101.45 mm^2; (A_{gt}-A_{ai})=2.27 mm^2;

3.4.3 Segmentation Accuracies

Using the CE-loss function, the mean accuracies for SDL/HDL models UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+ were **98.56±0.60, 98.19±0.83, 98.31±0.80, 98.44±0.70,** and **98.46±0.64,** respectively (Table 3.2, column 1). Similarly, using the DSC-loss function, the corresponding mean accuracies were **98.24±0.70, 98.29±0.74, 98.17±0.81, 98.13±0.73**, and **98.31±0.69**, respectively (Table 3.3, column 1).

Table 3.1 Hyper-parameters for experiments with SDL and HDL models.

Learning Parameters	UNet (CE)	UNet (DSC)	UNet+ (CE)	UNet+ (DSC)	SegNet (CE)	SegNet (DSC)	SegNet-UNet (CE)	SegNet-UNet (DSC)	SegNet-UNet+ (CE)	SegNet-UNet+ (DSC)
Optimizer	ADAM	ADAM	ADAM	ADAM	ADAM	ADAM	ADAM	ADAM	ADAM	ADAM
Partition Protocol	K10	K10	K10	K10	K10	K10	K10	K10	K10	K10
Training Data	786	786	786	786	786	786	786	786	786	786
Validation Data	87	87	87	87	87	87	87	87	87	87
Test Data	97	97	97	97	97	97	97	97	97	97
Initial Learning Rate	1e-4	1e-3	1e-3	1e-3	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
Max. Epochs	100	100	100	100	100	100	100	100	100	100
Mini Batch Size	6	5	5	10	5	10	10	5	10	10

Table 3.2 Results for five kinds of DL model using CE-loss function (Mean±SD).

Model	Accuracy	Sensitivity	Specificity	Precision	MCC
UNet (CE)	98.56 ±0.60	90.32±5.54	99.23±0.56	90.59±5.66	89.58±3.69
UNet+ (CE)	98.19±0.83	87.65±8.72	99.05±0.67	88.30 ±7.15	86.78±5.52
SegNet (CE)	98.31±0.80	89.41±7.13	99.04±0.64	88.32±6.53	87.81±4.93
SegNet-UNet(CE)	98.44±0.70	88.16±7.00	99.27 ± 0.45	90.69±5.14	88.46±4.42
SegNet-UNet+ (CE)	98.46±0.64	90.03±6.31	99.15±0.57	89.54±6.24	88.83±4.03
Mean±SD	**98.39**	**89.11**	**99.15**	**89.49**	**88.29**

Table 3.3 Results for five kinds of DL model using DSC-loss (Mean±SD).

Model	Accuracy	Sensitivity	Specificity	Precision	MCC
UNet (DSC)	98.24±0.70	95.25 ±4.47	98.45±0.74	86.93±4.70	86.50±4.51
UNet+ (DSC)	98.29±0.74	82.21±7.49	99.58±0.46	87.55±4.70	87.04±4.63
SegNet (DSC)	98.17±0.81	83.53±8.05	99.35±0.50	86.93±5.19	86.23±5.20
SegNet-UNet(DSC)	98.13±0.73	78.45±7.95	99.69±0.32	85.81±5.08	85.45±4.78
SegNet-UNet+ (DSC)	98.31±0.69	82.56±7.01	99.57±0.44	94.11±5.40	87.15±4.44
Mean±SD	**98.23**	**84.4**	**99.33**	**88.27**	**86.47**

3.5 Performance Evaluation

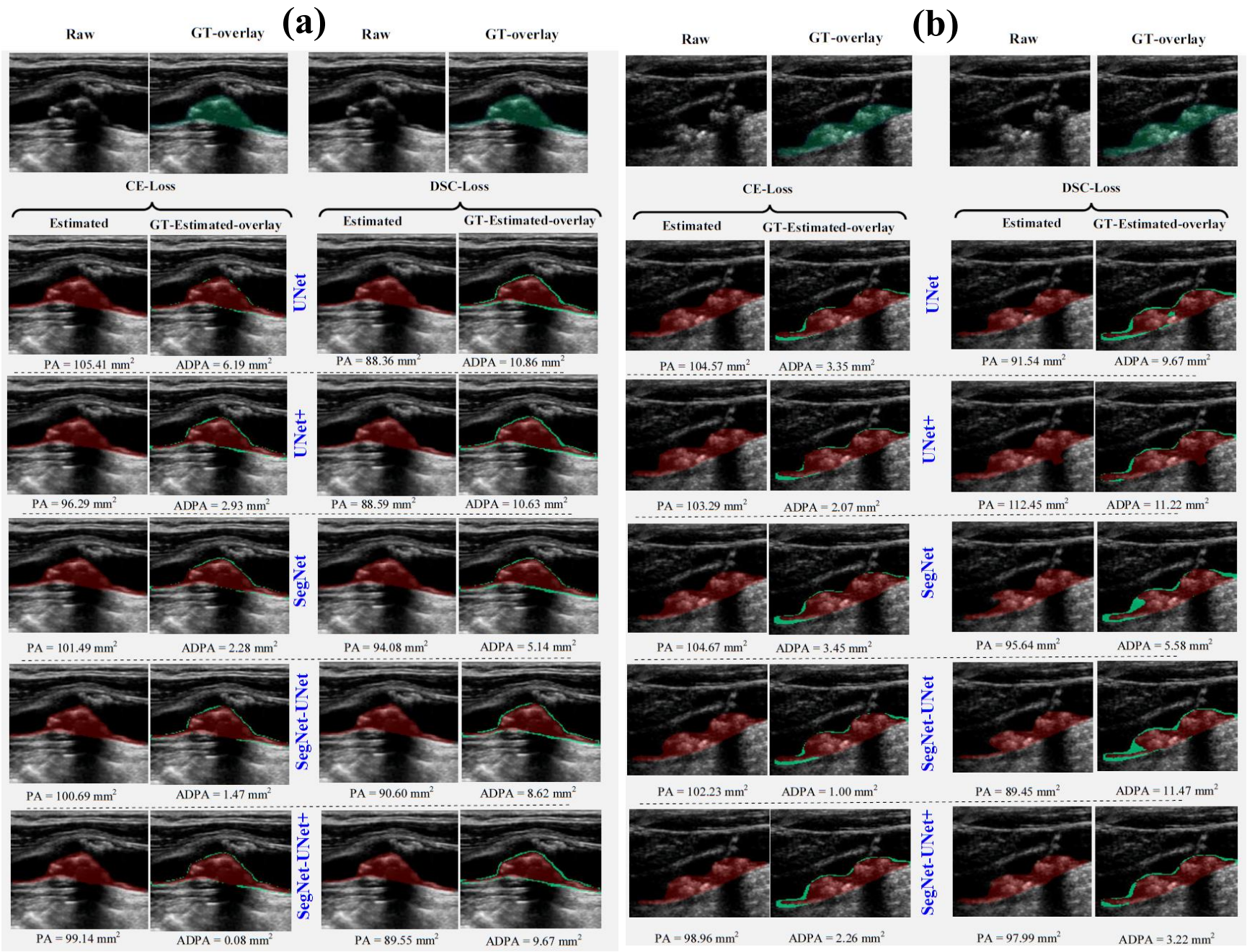
Performance evaluation is an important stage in CAD systems in which the detected output is compared against the gold standard input . A few essential strategies to evaluating the performance of the system are (i) regression curve and correlation coefficient between the estimated AI area and GT area, (ii) ROC curves and area under the curve computation given the threshold difference, and (iii) the figure of merit (FoM) of the AI systems. Tables 3.4 and 3.5 summarize the mean and SD values of Jaccard indices, Dice coefficients, and FoM for SDL and HDL models, respectively.

3.5.1 Jaccard Index for SDL/HDL models using CE-loss and DSC-loss function

The Jaccard index (JI) is also an index of similarity between GT- and AI-based outcomes, defined by Eq. 3.4:

$$\mathrm{JI} = \frac{\mathrm{TP}}{(\mathrm{TP+FP+FN})} \tag{3.4}$$

Using CE-loss function, the JI for the five models UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+ were **82.39±5.76**, **78.16±8.25**, **79.76±7.37**, **80.69±6.69**, and **81.20±6.23%**, respectively. For the corresponding DSC-loss function, the JI for the same models were **77.18±7.12**, **78.16±7.15**, **77.24±7.73**, **75.48±7.52**, and **78.34±6.85**, respectively.


(a)
(b)
Raw
GT-overlay
CE-Loss
DSC-Loss
Estimated
GT-Estimated-overlay
UNet
UNet+
SegNet
SegNet-UNet
SegNet-UNet+
PA = 105.41 mm²
ADPA = 6.19 mm²
PA = 88.36 mm²
ADPA = 10.86 mm²
PA = 96.29 mm²
ADPA = 2.93 mm²
PA = 88.59 mm²
ADPA = 10.63 mm²
PA = 101.49 mm²
ADPA = 2.28 mm²
PA = 94.08 mm²
ADPA = 5.14 mm²
PA = 100.69 mm²
ADPA = 1.47 mm²
PA = 90.60 mm²
ADPA = 8.62 mm²
PA = 99.14 mm²
ADPA = 0.08 mm²
PA = 89.55 mm²
ADPA = 9.67 mm²
PA = 104.57 mm²
ADPA = 3.35 mm²
PA = 91.54 mm²
ADPA = 9.67 mm²
PA = 103.29 mm²
ADPA = 2.07 mm²
PA = 112.45 mm²
ADPA = 11.22 mm²
PA = 104.67 mm²
ADPA = 3.45 mm²
PA = 95.64 mm²
ADPA = 5.58 mm²
PA = 102.23 mm²
ADPA = 1.00 mm²
PA = 89.45 mm²
ADPA = 11.47 mm²
PA = 98.96 mm²
ADPA = 2.26 mm²
PA = 97.99 mm²
ADPA = 3.22 mm²

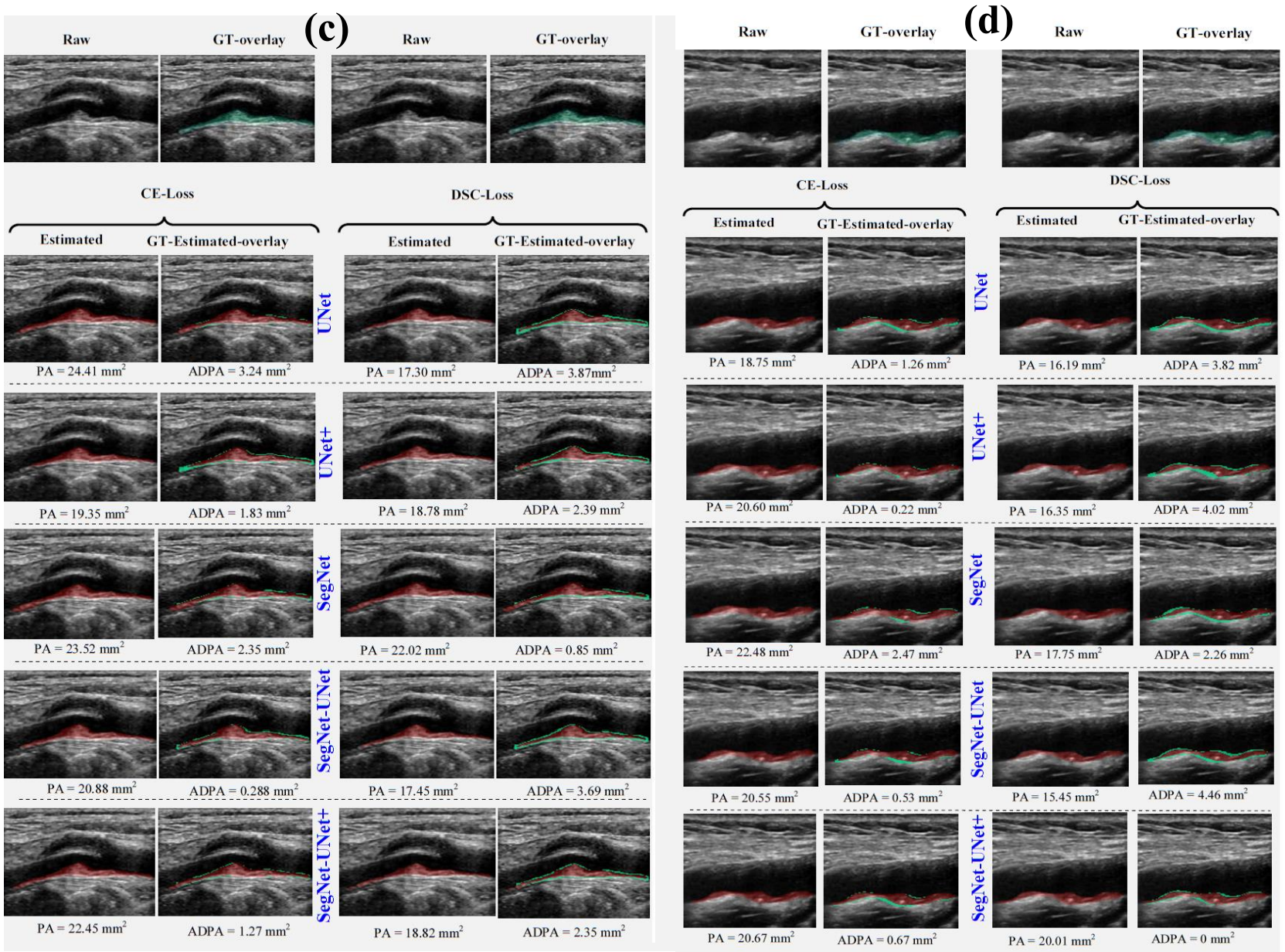


Figure 3.7 Estimated plaque area and difference area in (a), (b) (high-plaque), (c), (d) (low-plaque) images. Each (a), (b), (c), and (d) are divided into two columns consisting of CE-loss and DSC-loss. There are six rows: GT, UNet, UNet+, SegNet, SegNet-UNet, SegNet-UNet+.

Table 3.4 Jaccard *vs*. Dice performance for all five DL models using CE-loss (Mean±SD).

Model	Jaccard	Dice	**FoM**
UNet	82.39±5.76	90.23±3.59	99.85
UNet+	78.16±8.25	87.48±5.64	98.98
SegNet	79.76±7.37	88.54±4.86	99.19
SegNet-UNet	80.69±6.69	89.15±4.34	97.16
SegNet-UNet +	81.20±6.23	89.49±3.95	99.83
Mean±SD	**80.44±1.59**	**88.98±1.04**	**99.00±1.10**

Table 3.5 Jaccard *vs*. Dice performance for all five DL models using DSC-loss (Mean±SD).

Model	**Jaccard**	**Dice**	**FoM**
UNet	77.18±7.12	86.93±4.70	85.26
UNet+	78.16±7.15	87.55±4.70	87.78
SegNet	77.24±7.73	86.93±5.19	92.35
SegNet-UNet	75.48±7.52	85.81±5.08	83.25
SegNet-UNet+	78.34±6.85	87.68±4.48	88.10
Mean±SD	**77.28±1.13**	**86.98±0.74**	**87.35±3.43**

3.5.2 Dice Similarity Coefficient for SDL/HDL Using CE-loss and DSC-loss Function

Sorensen Dice index or Dice coefficient for segmentation is defined as Eq. 3.5:

$$\text{DSC} = \frac{2\times\text{TP}}{(2\times\text{TP})+\text{FP}+\text{FN}} \tag{3.5}$$

Using the CE-loss function, the DSC for the five models (UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+ models) are **90.23±3.59**, **87.48±5.64**, **88.54±4.86**, **89.15±4.34**, and **89.49±3.95**. Using the DSC-loss function, the corresponding five models gave **86.93±4.70**, **87.55±4.70**, **86.93±5.19**, **85.81±5.08**, and **87.68±4.48**. Figures 3.8 (a), (b), and (c) represent a bar chart showing mean Jaccard indices, mean Dice coefficients, and mean accuracy for the K10 protocol for all 10 models.

3.5.3 Figure of Merit

The figure of merit is defined based on the central tendency of the error. Let, $\mathcal{A}_{ai}(m,n)$ and $\mathcal{A}_{gt}(n)$ represent the plaque area using AI model '*m*' and GT, respectively, for the image '*n.*' Taking *N* as the total number of scans, the corresponding mean AI for model *m* and GT can be represented as $\bar{\mathcal{A}}_{ai}(m)$ and $\bar{\mathcal{A}}_{gt}$, [as defined in Eq. 3.3]. Then, the *FoM* can be expressed as Eq. 3.6:

$$FoM(m) = 100 - \left[\left(\frac{|\bar{\mathcal{A}}_{ai}(m) - \bar{\mathcal{A}}_{gt}|}{\bar{\mathcal{A}}_{gt}}\right) * 100\right] \tag{3.6}$$

Tables 3.4 and 3.5 represent FoM values for all 10 DL models. Column 3 represents the FoM for DL models with CE loss functions, whereas column 3 represents FoM for DL models with DSC-loss functions. It is evident from the table that each DL model with CE loss function performs better than the

DSC-loss function. UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+ with CE-loss have FoM values of 99.85, 98.98, 99.83, 97.16, and 99.19, respectively. Meanwhile, the same models with DSC-loss have FoM values of 85.26, 87.78, 88.10, 83.25, and 92.35. The means of all FoM values for DL with CE-loss is 99.00±1.10, which is greater than the mean FoM values for DL using the DSC-loss function 87.35±3.43. Figure 3.8 (d) is a bar chart representing the FoM values in increasing order.

3.5.4 Receiver operating characteristics analysis

Receiver operating characteristics (ROC) is a graphical representation of sensitivity and specificity, where a higher AUC confirms superior performance. MedCalc 18.2 software (from Belgium) was used to plot ROC curves for all 10 DL models (Figure 3.9). Also, the areas under the ROC curves are shown in Table 3.6, along with *p*-values. Using the CE-loss function, the AUC values for DL are 0.910, 0.911, 0.908, 0.905, and 0.898 for UNet, UNet+, SegNet, SegNet-UNet, and SegNet-UNet+, respectively. Using the DSC-loss function, the corresponding AUC for DL are 0.883, 0.889, 0.905, 0.889, and 0.907, respectively. In all five models, CE-loss-based DL models are either comparable or superior to DSC-loss-based DL models. Our observations show that the mean AUC of DL models using CE-loss is **1.2%** better than DSC-loss models.

Table 3.6 AUC (*p*-values) for all DL models using CE-loss *vs*. DSC-loss.

-	**CE-loss**	**DSC-loss**
Model	**AUC (*p*-value)**	**AUC (*p*-value)**
UNet	0.910 (<0.001)	0.883 (<0.001)
UNet+	0.911 (<0.001)	0.889 (<0.001)
SegNet	0.908 (<0.001)	0.905 (<0.001)
SegNet-UNet	0.905 (<0.001)	0.889 (<0.001)
SegNet-UNet+	0.898 (<0.001)	0.907 (<0.001)
Mean±SD	**0.906±0.005**	**0.895±0.011**

3.5.5 Correlation analysis between AI and GT

The regression curve is a handy and powerful statistical tool to establish a connection between two quantities. The correlation coefficient (CC) represents the extent to which the two quantities match. Its values lie within the range of 0 and 1. A higher CC value (close to 1) represents a high match between two quantities, while lower values represent a low level of matching. Figure 3.10 shows regression curves for all DL models. The left and the right columns represent the regression curves for DL with CE loss and DSC-loss functions. Table 3.7 represents the CC values for all five models and their mean and standard deviation. Mean±SD of CC values for DL for CE-loss models and DSC-models functions are 0.972±0.008 (p<0.001) and 0.978±0.004 (p<0.001), respectively.

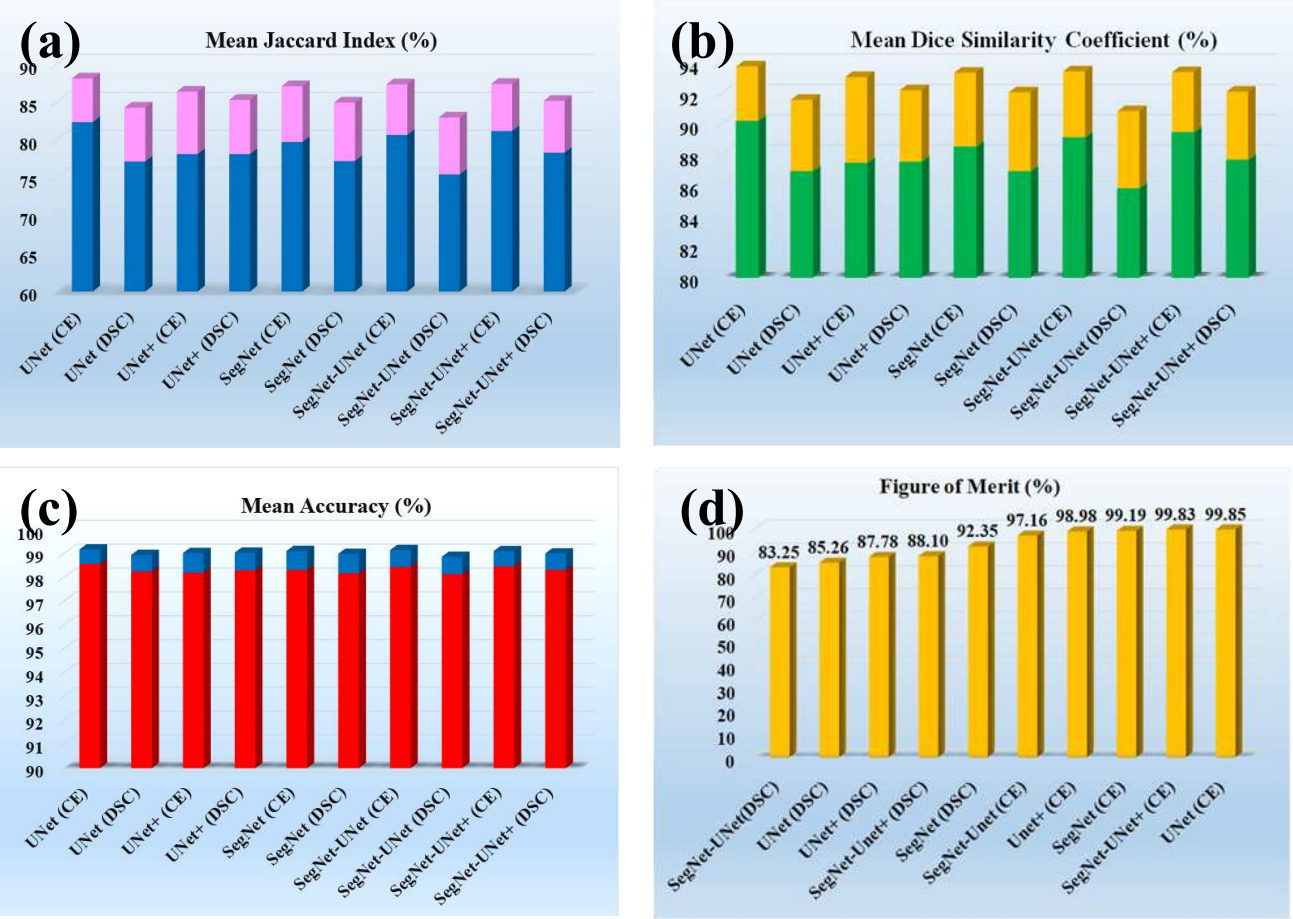


Figure 3.8 Bar charts showing mean performance parameters and FoM for ten DL models.

Table 3.7 Correlation coefficient between GT area and estimated area for all DL models.

Coefficient of Correlations between GT area and DL/HDL estimated area		
Model	**CE-loss**	**DSC-loss**
UNet	0.98 (p<0.001)	0.98 (p<0.001)
UNet +	0.96 (p<0.001)	0.98 (p<0.001)
SegNet	0.97 (p<0.001)	0.97 (p<0.001)
SegNet-UNet+	0.97 (p<0.001)	0.98 (p<0.001)
SegNet-UNet	0.98 (p<0.001)	0.98 (p<0.001)
Mean±SD	**0.972±0.008**	**0.978±0.004**

3.5.6 Bland-Altman Plots

Figure 3.11 shows Bland-Altman's (BA) plots for all 10 models. The Bland-Altman's plot shows the difference of the GT-area and AI-estimated area along the y-axis with the mean value of the GT-area and AI-estimated area along the x-axis. The spread of all points should be close to the middle line or within the ±1.96 SD. From Figure 3.11, it is clear that the spread within the BA plots in DL models with CE-loss is close to the middle line when compared to the DSC-loss model.

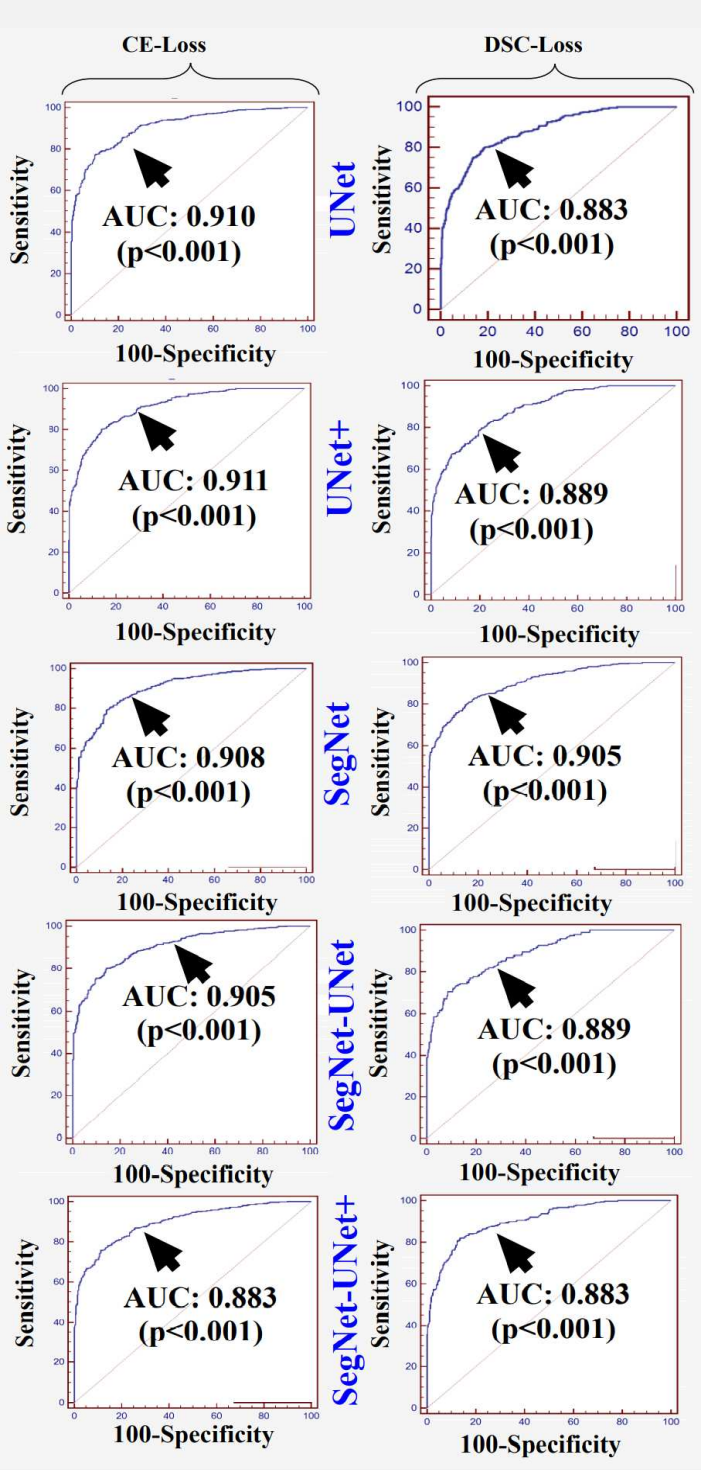


Figure 3.9 ROC curves for the sensitivity of the five models using CE-loss and DSC-loss.

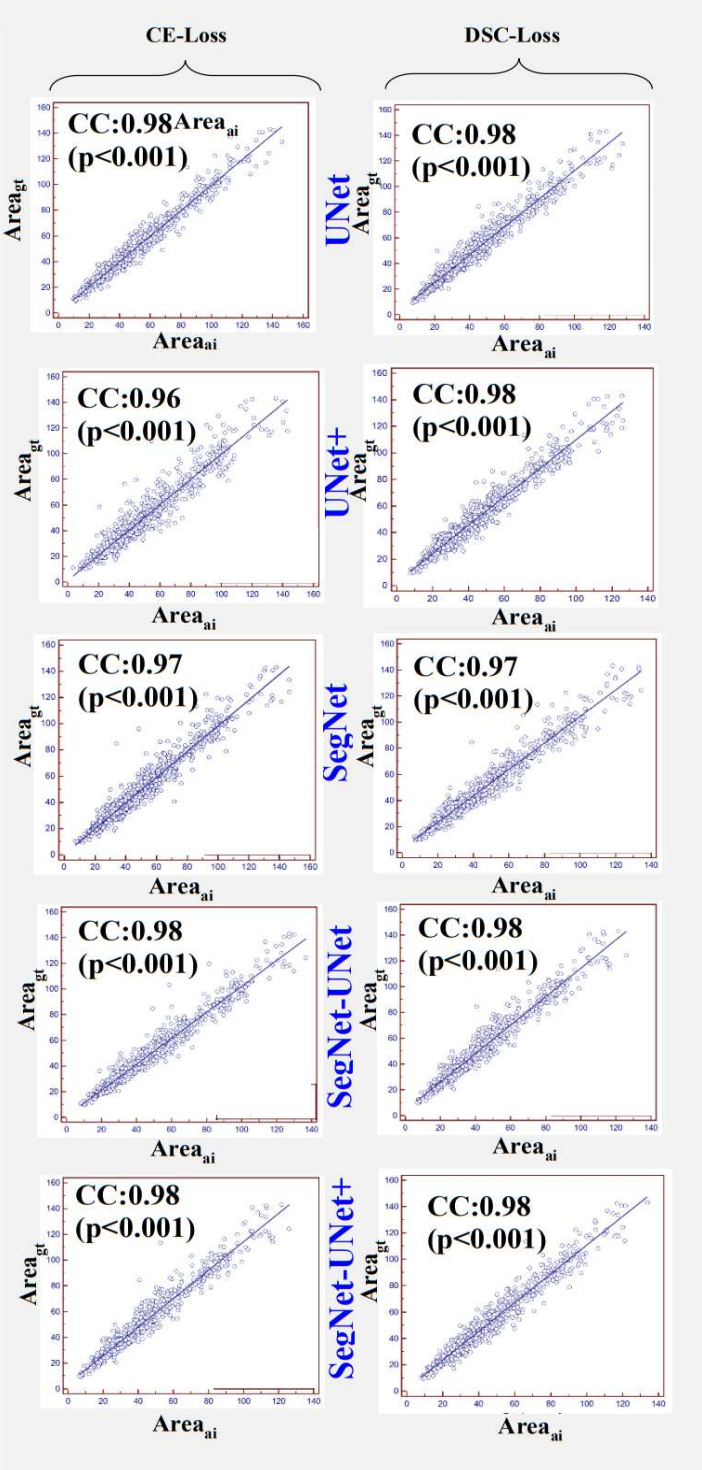


Figure 3.10 CC between GT area and AI-estimated area for five models using CE-loss and DSC-loss.

3.5.7 Cumulative distribution plot

Another way to choose a system's performance is to draw cumulative distribution curves. Figure 3.12 shows the cumulative distribution plot for SDL and HDL for CE-loss models (left column) and DSC-loss models (right column). All curves are plotted for several estimated images along the *y*-axis against the area error along the *x*-axis between 5 and 50 mm^2 at intervals of 5 mm^2. A dotted line connects the curve at 80% of the image database's corresponding area error.

Table 3.8 (a) summarizes models in ascending order of area error for 80% of the image database. As can be seen, 80% of the image database is < 6.2 mm^2 for the UNet model and 6 mm^2 for the SegNet-UNet model with CE-loss. The CDP demonstrates that DL models with CE-loss perform better than those with the DSC-loss model. Table 3.8 (b) shows the mean absolute difference (MAD) within the area between GT and AI-estimated models.

3.5.8 Sample size calculation

If we choose a sample size below the required number, the results may lag or lead due to the required number of samples. This is because the results can contain outliers and anomalies. When the sample size is too large with respect to the required sample size, the computational cost and time can increase—however, the results may be more accurate. Therefore, choosing the right sample size is crucial. To have the mathematical formula represent the sample size calculation, we used a *z*-score from the standard *z*-table as shown in Eq. 3.7 [169]:

$$N_p = (z)^2 \times \frac{\hat{p}(1-\hat{p})}{(MoE)^2} \qquad (3.7),$$

Where MoE is the margin of error, $\hat{p}$ represents data proportion. Using MoE confidence intervals of 4.45% or 95.54%, the standard *z*-score is 1.75. Therefore, with a 50% data proportion, our required sample size is 364. Thus, the database pool sample size is 166.5% greater than the number of samples required.

Table 3.8 Area error between AI and GT plaque for all DL models using CE and DSC-loss functions. (a) 80% of the database, (b) MAD.

	Area error (mm^2) - 80% of DB		MAD (mm^2)	
DL Model	**CE-loss**	**DSC-loss**	**CE-loss**	**DSC-loss**
SegNet-UNet	6.0	12.1	3.72	8.09
UNet	6.2	11.5	3.49	7.20
SegNet	7.0	8.2	4.21	4.97
SegNet-UNet+	7.8	9.8	4.21	6.21
UNet+	8.1	9.9	5.01	6.30
Mean±SD	**7.02±0.93**	**10.3±1.54**	**4.13±0.58**	**6.55±1.17**

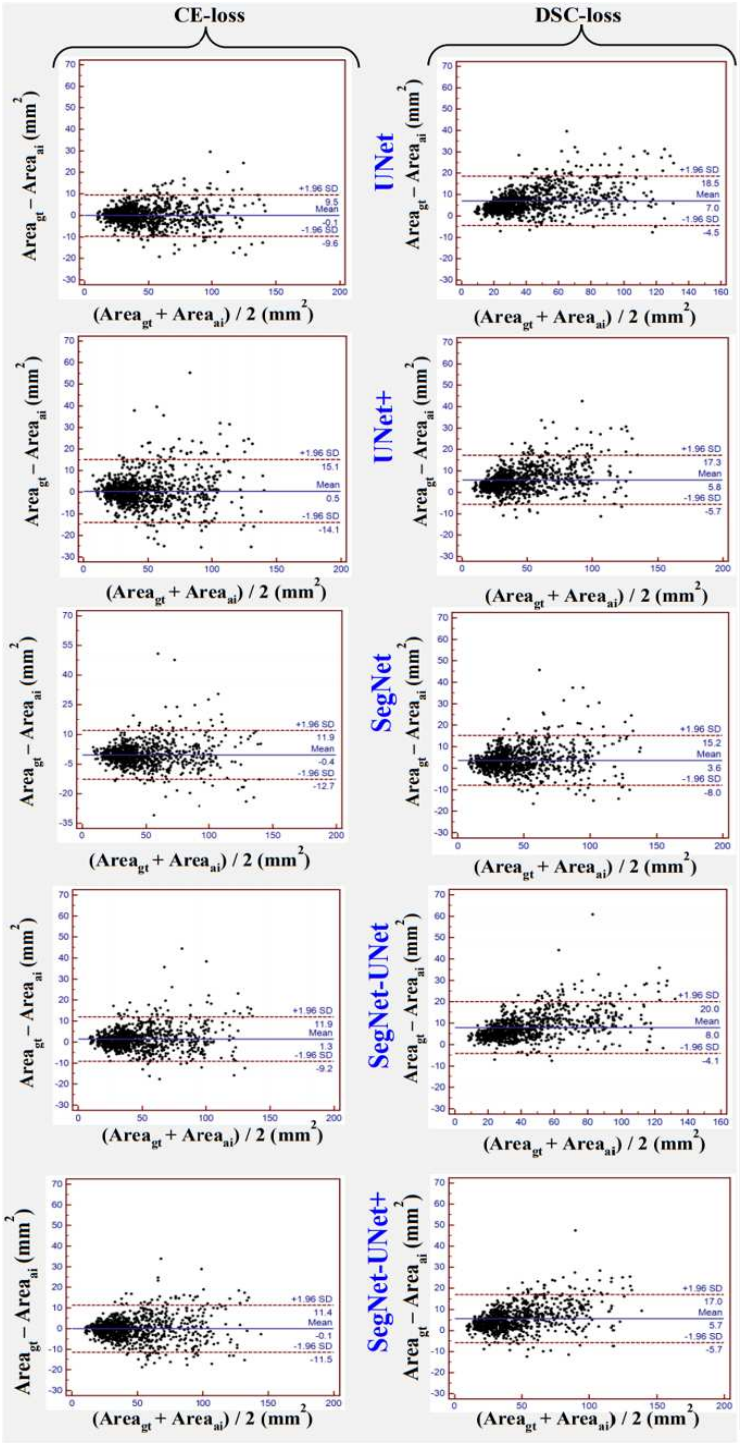


Figure 3.11 Bland-Altman plots for GT area and AI-estimated area for five models using CE-loss and DSC-loss.

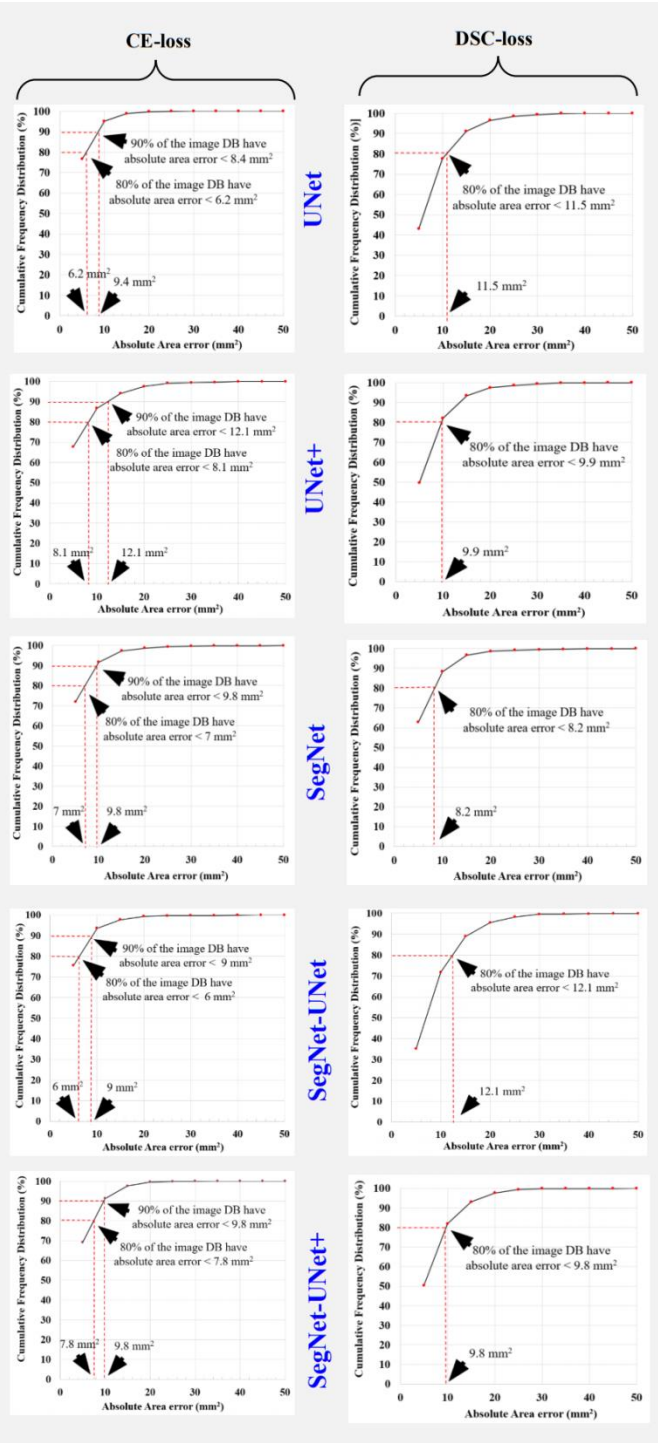


Figure 3.12 Cumulative distribution curves for area error between AI-estimated plaque area and GT plaque area for all DL models. Left: CE-loss. Right:

3.6 Discussion

The present study was the first of its kind for ICA plaque segmentation in B-mode carotid ultrasound scans. Ten DL models were implemented, of which six were SDL models (UNet (CE), UNet+ (CE), SegNet (CE), UNet (DSC), UNet+ (DSC), SegNet (DSC)) and four were HDL models (SegNet-UNet (CE), SegNet-UNet+ (CE), SegNet-UNet (DSC), SegNet-UNet+ (DSC)). Using a K10 cross-validation protocol using CE-loss framework—and keeping the threshold area difference of 10 mm^2 between the GT-area and AI-based model area—the models UNet (CE), UNet+ (CE), SegNet (CE), SegNet-UNet (CE), and SegNet-UNet+ (CE) showed performances with AUCs of **0.91**, **0.911**, **0.908**, **0.905**, and **0.898** (all having *p*-values of <0.001). Under same conditions and using DSC-loss framework, the UNet (DSC), UNet+ (DSC), SegNet (DSC), SegNet-UNet (DSC), and SegNet-UNet+ (DSC) models showed a performance with AUCs of **0.883**, **0.889**, **0.905**, **0.889**, and **0.907** (all having *p*-values of <0.001). Thus, the mean of AI-models using CE-loss functions shows an improved performance of **1.23%** (Table 3.6, attribute # 6) when compared to the mean of the AI-models using DSC-loss functions. The best performance improvement was **3.06%** (using UNet). The coefficient of correlation between DLPA-CE and GTPA were **0.98**, **0.96**, **0.97**, **0.98**, and **0.97** (all having *p*-values of <0.001), and between DLPA-DSC and GTPA were **0.98**, **0.98**, **0.97**, **0.98**, and **0.98** (all having *p*-values pf <0.001), respectively. Lastly, the online execution of the ICA segmentation system took less than one second.

The training system is offline and uses the GPU setting of the local machine (Lenovo, Intel Core i7, 8th generation processor, NVIDIA QUADRO, 8GB, 16GB RAM) that took about an hour for one fold in the K10 protocol for 90% training and 10% testing data (a total of 10 hours for the entire training of 873 US scans, with internal validation of 87 US scans for each epoch). Meanwhile, the online system took less than one second.

3.6.1 Benchmarking

Table 3.9 below is the benchmarking table, which has similar works performed by various researchers in previous years. Biswas *et al.* [107] used CNN and FCN methods (a class of AtheroEdge™ systems from AtheroPoint™, Roseville, CA) to find PA in ultrasound images. A pool of 250 CCA US scans was used in their study, demonstrating an average PA of 21.5794±7.9975 mm^2 and yielding an error of 2.7939±2.3702 mm^2 between the GT-area and AI-estimated model. Elisa *et al.* [43, 44] have proposed an FCN-based method for PA measurement. They used 396 CCA US scans and estimated PA for their database as 20.52 mm^2 and 19.44 mm^2 for two proposed DL networks.

Recently Zhou *et al.* [98] developed a UNet++ ensemble scheme using semi-automated plaque segmentation in CCA and ICA images for moderate to low- to high-risk plaque burden (<80 mm^2) patients (implemented on GPU NVIDIA Tesla P100 using Python 2.7 in Keras 2.2.4 framework).

Table 3.9 Benchmarking table in chronological order.

Authors	AST	AI Technique	Metric	#NP	PA or TPA
Biswas *et al.* [170]	CCA	CNN, FCN	cIMT, TPA	250	Mean PA: 21.5794±7.9975mm^2 PA error: 2.7939±2.3702 mm^2
Elisa *et al.* [43]	CCA	FCN	TPA	396	20.52 mm^2, 19.44 mm^2
Zhou *et al.* [98]	CCA ICA	UNet++ ensemble	PA, TPA	144	Best TPA error: 5.55±4.34 mm^2
Proposed 2021	ICA	UNet(CE) UNet (DSC) UNet+(CE) UNet+(DSC) SegNet(CE) SegNet(DSC) SegNet-UNet (CE) SegNet-UNet (DSC) SegNet-UNet+ (CE) SegNet-UNet+ (DSC)	PA	970	Best PA error: 3.49 mm^2 using UNet with CE-loss

AST: Artery Segment Type; NP: # of patients; PA: Plaque Area; TPA: Total Plaque Area; CNN: Convolutional Neural Network; FCN: Fully Convolutional Network; CE: Cross-Entropy; DSC: Dice Similarity Coefficient.

The primary architecture (UNet++) was applied to a single dataset of 144 patients, where three partitions were created, consisting of three kinds of training data (such as 33, 33, 34) and test data consisting of 44 patients. These three partitions were used for the ensemble model. The UNet++ system reported a DSC of 83.3-85.7%. In comparison to Zhou *et al.* 's study, we proposed four fully-automated (having no manual interaction for creating the region-of-interest) HDL and six fully-automated SDL for high- to moderate- to high-risk plaque burdens (<143 mm^2) for segmentation in ICA US scans. Our system was implemented on GPU NVIDIA™ Quadro P4000 8GB using MATLAB™ 2020a. Our performance of HDL was comparable to SDL. However, CE-loss was superior in all 10 DL models, and the best DSC was 90.23±3.59% (UNet with CE-loss). Further, our online system was <1 sec per scan.

3.6.2 A Note on SDL and HDL

The HDL models are a concatenation of two SDL models. These models have the fundamental properties of network depth (stages). For each new application, there is a unique SDL model. As a result, the SDL model requires retuning or retraining to run the SDL for other applications. Retraining and retuning are time-consuming. On the contrary, in the HDL models, if the depths of the two SDL models are different, retraining and retuning are not required. As a result, HDL is more advantageous than SDL. Within the context of this advantage, our four HDL models (SegNet-UNet (DSC) SegNet-UNet (CE)) are superior to the SDL model. Theoretically, HDL is more flexible and powerful than SDL models. However, more is needed to provide a deeper evaluation.

3.6.3 A Special Note on CE Vs. DSC

An interesting note about these 10 architectures is the loss function, the so-called heart of the system. We adapted both CE-loss and DSC-loss functions. Our observations showed that architectures using the CE-loss function consistently demonstrated higher performance than architectures with DSC-loss functions. This means that in terms of all AI attributes—such as segmentation accuracy, sensitivity, specificity, precision, MCC, AUC, JI, DSC, FoM, and CC, the difference in the GT-area and AI-estimated area, sensitivity—the CE-loss model performed better than or comparably to DSC-loss models (see Table 3.10). This implied that UNet (CE) > UNet (DSC), UNet+(CE)>UNet+(DSC), SegNet (CE)>SegNet(DSC), and SegNet-UNet(CE)>SegNet-UNet(DSC) and that SegNet-UNet+(CE) > SegNet-UNet+(DSC), where ">" represents higher performance.

Table 3.10. Performance of the mean of five* AI models: CE-loss vs. DSC-loss.

SN	PE-metrics	CE-loss	DSC-loss	Improvement (%)	Agreement
1	Accuracy (%)	98.39	98.23	0.17	✓
2	Sensitivity (%)	89.11	84.40	5.59	✓
3	Specificity (%)	99.15	99.33	-0.18	~
4	Precision (%)	89.49	88.27	1.38	✓
5	MCC (%)	88.29	86.47	2.10	✓
6	AUC (0-1)	0.906	0.895	1.23	✓
7	Jaccard (%)	80.44	77.28	4.09	✓
8	Dice (%)	88.98	86.98	2.30	✓
9	FoM (%)	99	87.35	13.34	✓
10	CC ($\bar{\mathcal{A}}_{ai} vs. \bar{\mathcal{A}}_{gt}$)	0.972	0.978	-0.61	~
11	MAD	4.13	6.55	36.95	✓

* UNet, UNet+, SegNet, SegNet-UNet, SegNet-UNet+;✓ CE-loss>DSC-loss; ~ Comparable

Note that CE loss is defined as the logarithmic loss or log loss. Eq. 3.1 defines the CE-loss, where y_i is the input GT label '1', 1-y_i is the GT label '0', and a_i is the SoftMax classifier probability. In the segmentation case, class values '0' and '1' from the predicted pixel class are compared against the GT pixel class. An overall score 'a_i' is calculated for the segmented image and expresses how close the segmented image is to the GT image. This overall score is a logarithmic calculation that yields a high score for large differences (close to 1) and a small score for small differences (close to 0). Finally, we aim to minimize CE-loss during training, which is lower in CE loss—thus, it is a better segmentation model.

Since ultrasound images are fuzzy and have very little information other than plaque area, the overall score (SoftMax classifier probability) is high for large differences. Further, the ADAM optimization technique reduces the loss with rising epochs. In the case of DSC-loss, the loss is calculated between the AI-based and GT pixel values based on the Dice coefficient match. However, DSC-loss performed better in some other general images, but since ICA images are noisy and fuzzy, the loss due to low Dice similarity between the GT and predicted class is higher than CE-loss.

3.6.4 The Trade-Off Between The Interval Gap During Data Preparation And Augmentation

Data collection is a very expensive, time-consuming, and complex process in medical image analysis. ICA is not commonly scanned on all diagnostic centres. Hence, we faced challenges while obtaining the ICA US scans with limited volumes, totalling 100 patients. After excluding erroneous data, we were left with only 970 US scans. DL techniques are known for exhausting lots of data for training. Therefore, we converted the US videos (movies) into still images using an interval of 10 between consecutive image frames of the patient. It is evident that US images have a specific difference in plaque profile during US scans—therefore, each selected image frame can be treated as a new independent US scan. However, our system can be considered as biased towards the database. Therefore, we reduced the gap between two consecutive image frames to 0, meaning we used only 97 images (1X) for training and testing. By using these 97 US scans (interval gap = 0) and running the K10 cross-validation system, we obtained a mean accuracy of **95.7%** and a DSC of **71.04%**.

3.6.5 Augmentation Protocol

There are certain other techniques by which a single image can be used multiple times via the image augmentation method [150]. We used rotation (-5^0 to $+10^0$), reflection, and translation (20 pixels) methods to augment the image database. Thus, we prepared 1.5X (using 0.75X original + 0.75X rotation), 2X (using original + rotation), 2.5X (using 0.8X original + 0.8X rotation + 0.8X reflection), and 3X (using original + rotation + reflection) [142, 143]. However, we achieved the highest accuracy (98.34%), Dice coefficient (88.52%), and Jaccard index (80.28%) for the 2X database, which is comparable to our 10X database experiment. Still, the amount of crispiness in segmentation is higher with a large database.

The pyramid depicted in Figure 3.13 shows the evidence of segmentation example using *augmentation protocol* and using 1X, 2X, and 3X databases and UNet with CE-loss. The first row of Figure 3.13 shows the segmentation using UNet (CE) in the 1X database, which has a poor plaque segmentation region labelled "*under-segmentation*." The second row in the same figure shows the segmentation mask for the 2X database, which is considered the "*best segmentation*," whereas the segmentation results using the 3X database is "*over-segmentation,*" as shown in the third row.

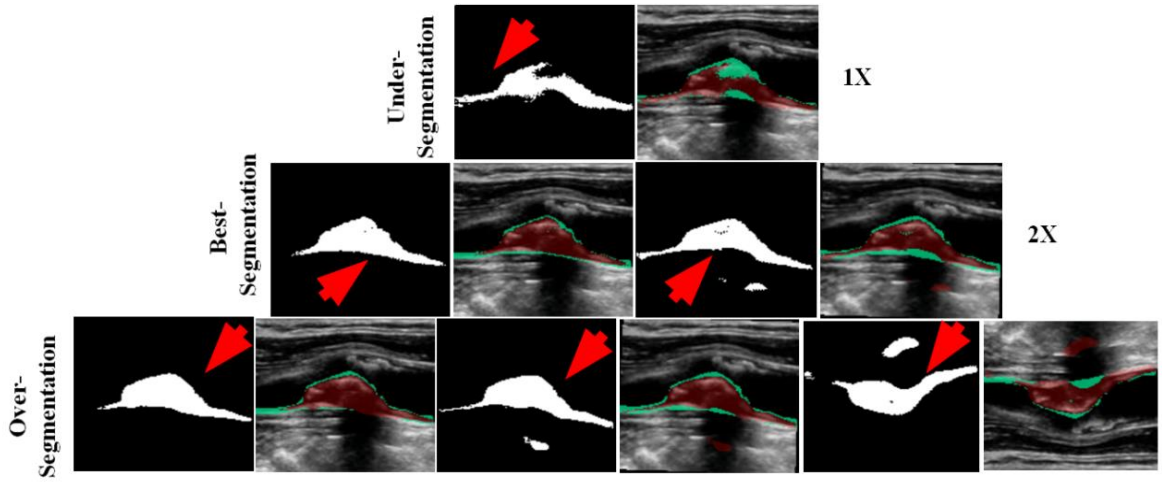


Figure 3.13 Segmentation of images in 1X, 2X, and 3X databases using the UNet (CE) model.

Figures 3.14 (a) and (b) show the accuracy results of segmentation in 1X, 1.5X, 2X, 2.5X, and 3X databases using CE-loss and DSC-loss, respectively. It is clear from the black arrow (pointing 8 o'clock) in Figures 3.14 (a) and 14 (b) that the 2X (original+rotation) database gives the *best segmentation* results from the entire pool of augmentation databases. This confirms that segmentation results using 10 intervals are comparable to those using no intervals with a 2X augmentation protocol.

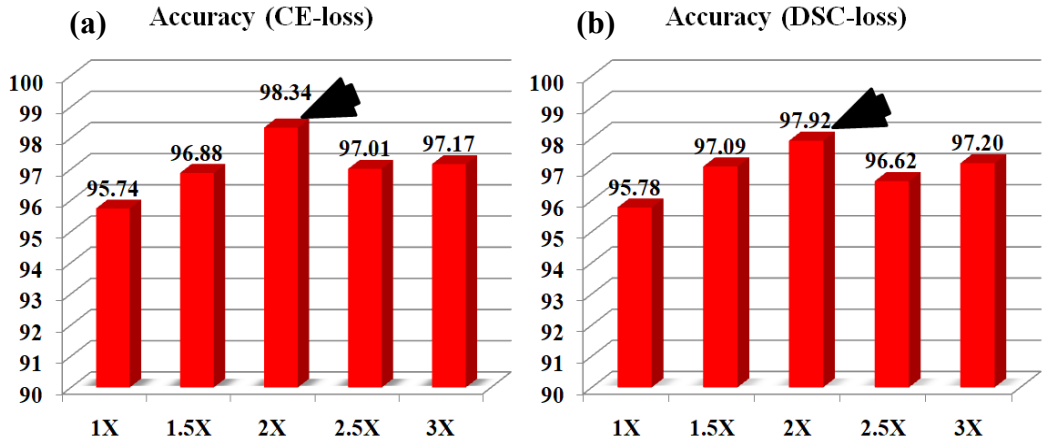


Figure 3.14 Optimized accuracy for augmented databases using (a) CE-loss and (b) DSC-loss. The best accuracy was obtained using the augmentation protocol for 2X.

3.6.6 Identification of plaques at risk of rupture

There are several challenges in the identification of rupture-prone plaque or plaque at the risk of rupture. This is due to the variability in the instrument's US gain and variability in plaque tissues. The symptomatic plaques tend to have a large hypoechoic juxta-luminal black area close to the lumen and exhibiting significant stenosis. However, there is no bright echogenic cap [171], and, thus, it is hard to

visualize. Contrarily, asymptomatic plaques show hyperechoic regions with white intensities reflecting stable plaque [122, 172], typically calcium. Thus, overall, the characterization of plaque tissue types viewed by the naked eye is challenging and offers inter-observer variability [173]. The above claim has been validated against histology [172].

Due to the above reasons, vulnerable plaques are difficult to identify and automatically delineate. Therefore previous methods were manually characterized [142, 143, 174]. Such plaque structures are also irregular and have more considerable variations in plaque distribution, thus making it tedious to delineate. Since radiologists (sonographers, or vascular surgeons) do not have enough time during the screening process to accurately delineate them and quantify the plaque area or volume, it is even more challenging. Thus, such plaques are missed during the routine screening. Further, internal carotid artery (ICA) is not always adapted during screening in all diagnostic centres, especially in developing countries or underdeveloped regions in advanced nations due to a lack of radiological resources, primarily because of the cost and time required.

AI becomes very useful in such situations [93, 133, 170]. Since the AI system is trained to identify ICA plaques, these trained models can be used online for the automated delineation of the plaque regions in US patient's scans [107]. Previously, such cross-validation approaches have been very promising in US applications [142, 143, 175], and such cross-validation approaches have been very promising in ultrasound and non-US frameworks [176]. Since our system takes less<1 second per image, it is practical to introduce such an AI-based system to detect rupture-prone plaques (or vulnerable plaque detection). Also, since AtheroPoint™ has utilizes the AtheroEdge™ system for diagnostic centres [44, 177, 178], these hybrid DL models can be integrated into their pipelines to detect all categories' plaque automatically.

3.6.7 Plaque detection in positive remodelling cases

Positive remodelling is defined as an outward thickening of the vessel wall with and without restricting the lumen region [179]. Our group has previously conducted studies with analyses demonstrating PR cases [74]. Our ICA segmentation study observed moderate to high positive modelling as shown in Figures 3.15 (a), (b), (c), and (d). Further, we noticed that the lumen region was also restricted by 40% to 50% (as mentioned in the demographics section of our research). It should be emphasized that in our HDL model, the system accurately identified the plaque region in the positive remodelled cases. This means the HDL model closely followed the ground truth (GT) positive remodelling scenario. The difference between the estimated HDL positive remodelled plaque region (red) and GT positive remodelled region (green) is very small, ensuring the higher accuracy of the HDL model in a positive remodelled zone. This is shown as the overlay of red (HDL) and green colours (GT) with

respect to the grayscale background region. Thus, we conclude that all the models can handle the positive remodelled cases of our ICA database.

3.6.8 Strengths/Weakness and Extensions

The main strength of the overall system is the power of the DL/HDL systems to learn the behaviour of ICA regions using CE-loss or DSC-loss functions. Also, the system is generalized using the K10 protocol. Further, the overall system takes less than one second for a test ICA scan. HDL models such as SegNet-UNet had the best-performing architecture. The system utilized the GPU resource.

The image quality of our data was highly variable and low in resolution. An older generation (U22) model of Philips Ultrasound was used. We believe having a higher image quality would improve the performance further. This can be accomplished using speckle reduction methods [180]. The system can be extended to other medical imaging segmentation paradigms. By using the big data paradigm [181], larger cohorts can be considered. In the future, ICA stenosis can be computed in the DL framework [133] along with the variability analysis as previously conducted by our team [30, 31].

This was a pilot study, and only the K10 protocol was used, which can be extended to other sets of cross-validation protocols in the future [182]. The automated plaque area can also be evaluated against the maximum plaque height [183]. Further, to achieve a solid benchmarking system, better machine learning models, such as ensemble or multiclass paradigms, can be designed. Lastly, one can use PTC approaches for plaque tissue characterization [36] on these segmented plaque regions.

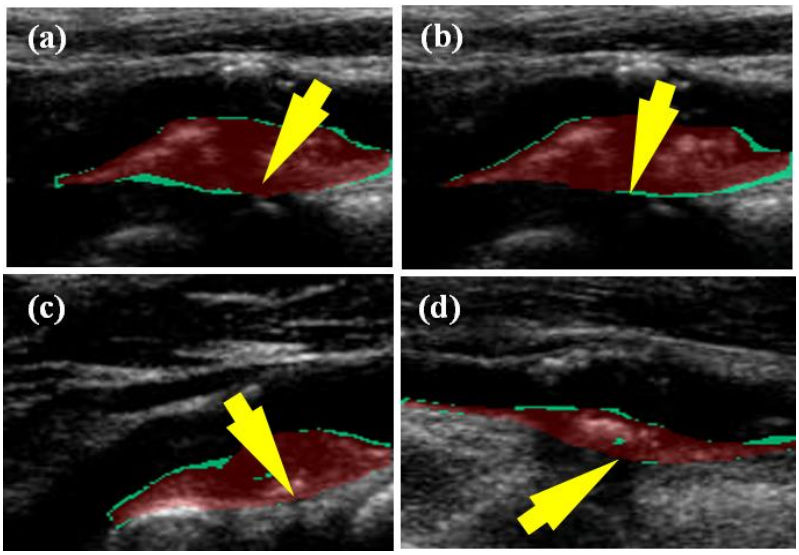


Figure 3.15 Examples showing positive remodelling.

3.7 Conclusion

Under the class of AtheroEdge™ systems from AtheroPoint™, Roseville, CA, this study was the first of its kind to segment carotid plaque in B-model ICA US scans. Ten DL models were implemented, of which six were solo DL models and four were HDL models. We validated our hypotheses that solo DL and HDL models are comparable for plaque segmentation in ICA and that the CE-loss function is superior to the DSC-loss function. The selection of 10 intervals in US videos was justified using an augmentation protocol. SegNet-UNet was the best-performing hybrid architecture, with a segmentation accuracy of **98.44%**, demonstrating 6 mm^2 of area error for 80% of scans with an AUC of **0.905** ($p<0.001$) using a 10% threshold. Overall, the online AtheroEdge™ 3.0 system performed segmentation in less than one second. The segmentation system can be applied to other arterial imaging techniques as part of future work.

3.8 Proposed Extension for Next Chapter

Current work focuses only one segment of the carotid artery i.e. ICA section. Since all symptomatic ICA images have moderate to high plaque content, results are quite impressive. However, CCA images have low-to moderate plaque deposition. Therefore, the system can be extended to segment the plaque from common carotid artery section. Further, we need to validate our system against a benchmark system such as AtheroEdge 2.0. In the next chapter we have incorporated these changes.